# Bayesian Hierarchical Models to Augment the Mediterranean Forecast System

Mark Berliner
Department of Statistics
Ohio State University
1958 Neil Avenue, Columbus, OH, USA 43210-1247
phone: (614) 292-0291 fax: (614) 292-2096 email: mb@stat.ohio-state.edu

Emanuele Di Lorenzo
School of Earth and Atmospheric Sciences
Georgia Institute of Technology
311 Ferst Drive, Atlanta, GA, USA 30332-0340
phone: (404) 894-3994 fax: (404) 894-5638 email: edl@eas.gatech.edu

Ralph F. Milliff
Colorado Research Associates Division
NorthWest Research Associates Inc.
3380 Mitchell Lane, Boulder, CO, USA 80301-5410
phone: (303) 415-9701 x216 fax: (303) 415-9702 email: milliff@cora.nwra.com

Christopher K. Wikle
Department of Statistics
University of Missouri-Columbia
146 Middlebush, Columbia, MO, USA 65211
phone: (573) 882-9659 fax: (573) 884-5524 email: wikle@stat.missouri.edu

Award Number: N00014-07-10512
http://www.cora.nwra.com/MedBhm (password protected)

## LONG-TERM GOALS

Year 2 of Phase 1 "Bayesian Hierarchical Models (BHM) to Augment the Mediterranean Forecast System ( MFS)" ended in May 2007. Long-term goals for Phase I included: a) development of an ensemble ocean forecast methodology based on a surface wind BHM (MFS-Wind-BHM) in data assimilation and forecast steps of the MFS; and b) development of a BHM for time-dependent background error covariance evolution (MFS-Error-BHM) in the MFS data assimilation system .

Phase II of the project was initiated in June 2007. Long term goals for the second phase include the development of a BHM to guide ocean model super-ensemble experiments, in both multi-model and the multi-parameter experimental designs. The MFS ocean forecast model will be modified for multi-parameter super-ensemble experiments, and MFS will be joined by a Mediterranean Sea implementation of the Regional Ocean Modeling System (MedROMS: http://www.med-roms.org) in multi-model super-ensemble experiments.

| **Report Documentation Page** | *Form Approved* *OMB No. 0704-0188* |
|---|---|



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **30 SEP 2007** | **Annual** | **00-00-2007 to 00-00-2007** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**Bayesian Hierarchical Models To Augment The Mediterranean Forecast System** | 5a. CONTRACT NUMBER<br>5b. GRANT NUMBER<br>5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER<br>5e. TASK NUMBER<br>5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Ohio State University,Department of Statistics,1958 Neil Avenue,Columbus,OH,43210** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**code 1 only**

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **11** | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

## OBJECTIVES

Research objectives for the completion of Phase I in the past year included:

1. developing metrics to distinguish versions of the MFS-Wind-BHM that differed in the level of explicit physics in the prior distribution models;

2. documenting the versions and distinctions of MFS-Wind-BHM, with a goal to identify the model version to be implemented in MFS operations;

3. developing the MFS-Error-BHM for testing in the MFS research ocean model; and

4. drafting manuscripts to document MFS-Wind-BHM and MFS-Error-BHM developments.

Research objectives in the organization of Phase II include:

1. run-matrix design for multi-parameter super-ensemble experiments in MFS; and

2. Mediterranean forecast model development plans for MedROMS;

## APPROACH

Probability models, using the Bayesian Hierarchical Model (BHM) formalism, are being adapted to practical ocean forecast issues in the Mediterranean Forecast System (MFS). The first two-year project explored means of generating MFS ensemble initial conditions and forecasts from realizations of the surface wind forcing. Also, a general method was developed to introduce time-dependence in background error covariance specifications for optimal interpolation data assimilation models. The BHM formalism will be extended to Super-Ensemble ocean forecast issues in the Phase II of the project (succeeding two years).

## WORK COMPLETED

### *MFS-Wind-BHM*

Four related versions of MFS-Wind-BHM were developed as we explored issues of explicit vs. implicit physics in models for the prior distribution, and the necessity of a fine-scale spectral constraint (Table 1). In order to evaluate and compare these models, we developed a set of 4 metrics, including: 1) examining posterior distributions of the random coefficients that lead each term in the dynamical model for the prior distribution (i.e. the process model stage); 2) comparing sea-level pressure (SLP) posterior mean fields with ECMWF analyses for snapshots during the forecast period; 3) examining kinetic energy vs. wavenumber (KE vs. k) spectra for zonally oriented spatial series that overlap cross-swath lines from the QuikSCAT data and zonal lines from the ECMWF analyses; and 4) observing the spread generated in ocean forecast model initial fields forced (during the data assimilation stage) by realizations from MFS-Wind-BHM posterior distribution.

***Table 1. Versions v5 – v8 of MFS-Wind-BHM distinguished by the forms of process models for the prior distribution. The v5 model is very similar to the model developed by Royle et al. (1998). The v6 model includes explicit terms in the process model for time-dependent terms that arise from the Rayleigh Friction Model equations. Models v7 and v8 are the counterparts of v6 and v5, respectively, and they include a small-scale nested wavelet basis set to enforce spectral slope (as developed by Wikle et al. 2001).***

| ***MFS-Wind-BHM version*** | ***Explicit time-dependent terms*** | ***Wavelets*** |
|---|---|---|
| v5 | no | no |
| v6 | yes | no |
| v7 | yes | yes |
| v8 | no | yes |

Examples of the first three of these metrics for the v7 MFS-Wind-BHM are provided in Figures 1-3 below. Sample initial condition spreads were provided for sea-surface height and sea-surface temperature in the annual report from last year.

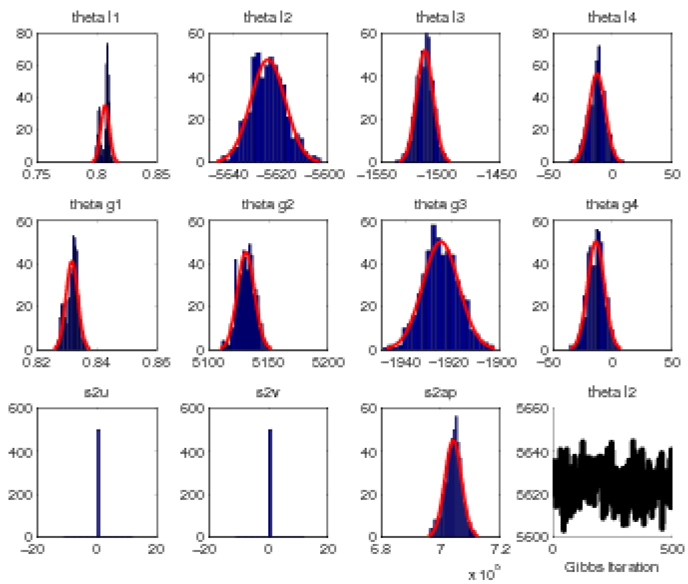


***Figure 1. Posterior distributions for random coefficients in the explicit time-dependence /wavelet model (v7). Row 1 contains the coefficients for the u equation. The coefficient $\theta_{l2}$ multiplies the geostrophic term; and $\theta_{l3}$ multiplies the ageostrophic term (i.e. the down-gradient, pressure gradient term). These terms dominate the momentum balance for u. A similar balance pertains for the v equation in row 2. Magnitudes for the coefficients leading explicit time dependent terms are not as large as the geostrophic and ageostrophic coefficients. The third row depicts posterior variance estimates for u, v, and SLP. The $\theta_{l2}$ values for each Gibbs iteration are shown in the lower right panel.***

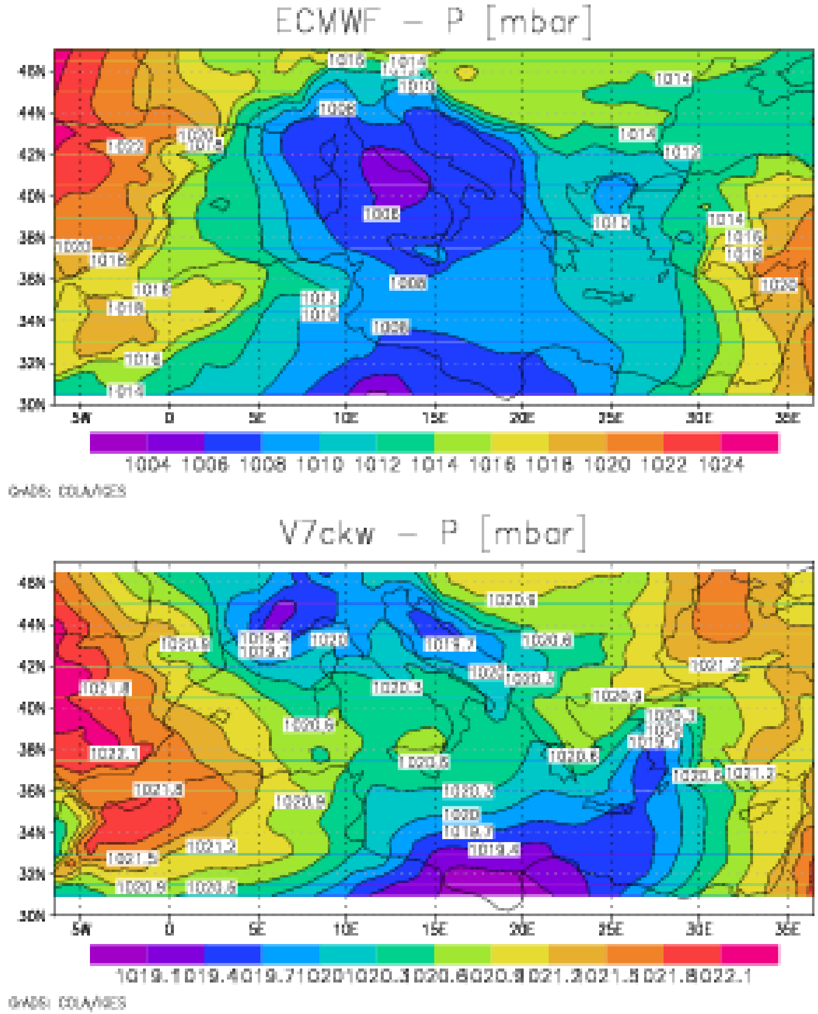


***Figure 2. Sea-level pressure (SLP) comparison ECMWF analysis (top) vs. MFS-Wind-BHM posterior mean (bottom) for a snapshot during the forecast period. The ECMWF analysis field does not influence the BHM during the forecast period. While large-scale features are well-reproduced, the details of the SLP field in the vicinity of the low pressure center over western Italy are not similar.***

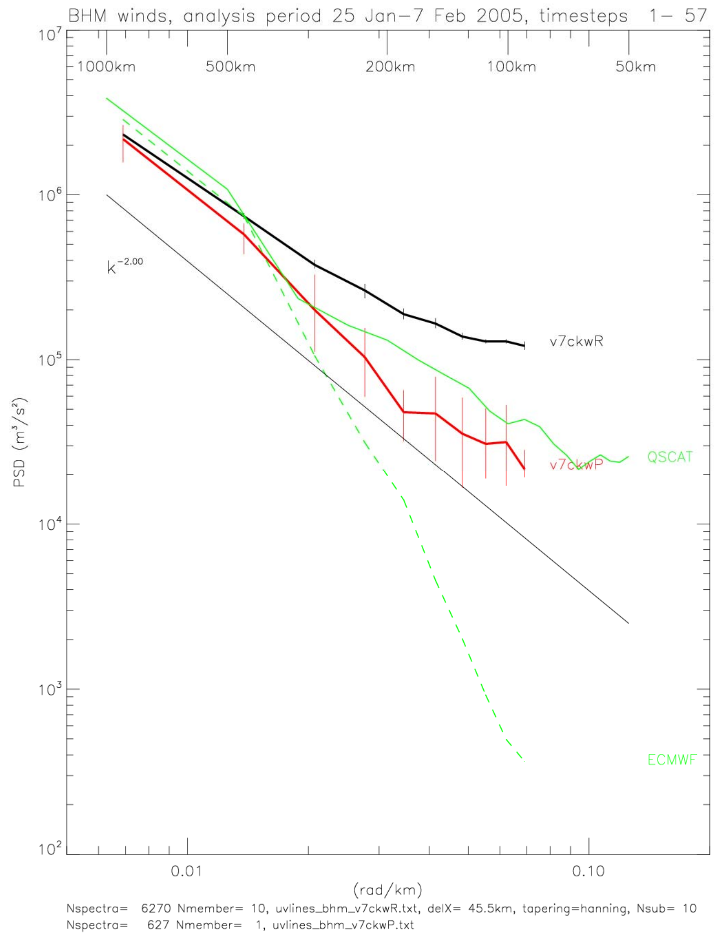


***Figure 3. Kinetic energy vs. wavenumber spectra for zonal wind spatial series from zonal lines in the Mediterranean Sea for the analysis period 25 January – 7 February 2005. Green spectra are for QuikSCAT data (solid line) and the ECMWF analyses (dashed line). A reference curve representing a $k^{-2}$ power law is indicated by the dashed black line, closely matching the slope for the QuikSCAT data. The average spectrum for 10 realization s from the posterior distribution for MFS-Wind-BHM is the solid black (bold) line, and the posterior mean spectrum is drawn in solid red (bold). The spectral spread computed from 10 different estimates of the posterior mean is indicated in vertical red lines, the spread in the 10 realizations from the posterior distribution (smaller range) is indicated by vertical black lines. The black and red spectra are sufficiently close to the QuikSCAT spectrum for this period.***

The v7 and v8 models are being considered for operational implementation in MFS. Nested wavelet basis functions are required to enforce high-wavenumber spectral properties that are a resilient feature of the QuikSCAT data, missing entirely from numerical weather prediction winds. Calculations are underway now to enable longer term (i.e. several year) spectral comparisons with QuikSCAT. The v7 model SLP comparison (Fig. 2) is not sufficient to select this model outright, so v8 is being considered. Manuscripts describing details of MFS-Wind-BHM development (Milliff et al, 2007), and impacts on the MFS forecasts (Bonazzi et al, 2007) are in preparation.

*MFS-Error-BHM*

Two forms of data comprise data stage distributions (likelihoods) in MFS-Error-BHM. They are: 1) the observation/forecast *misfits,* **d**, obtained principally from ARGO float profiles; and 2) the forecast difference with respect to a long-term average forecast for the same year-day (so-called *anomalies*, **q**). The misfits are given by $\mathbf{d} = \mathbf{H}\mathbf{x}^{f} - \mathbf{x}^{ARGO}$, where **H** is the observation operator that moves the model forecast $\mathbf{x}^{f}$ to the ARGO location for each observation time. The anomalies are given by $\mathbf{q} = \mathbf{x}^{f} - \mathbf{x}^{avg}$.

The data stage distributions are combined with a process model (prior) for the time-dependent error process, **e,** as described in the annual report from last year. Depth vs. time profiles of the mean of the posterior distribution for **e** are shown in Fig 4. for one of 13 Mediterranean sub-regions (Gulf of Lions) treated in MFS. The error process in temperature is concentrated above the thermocline, and exhibits a seasonal signal. The error process for salinity also exhibits some seasonality. It is also concentrated in the upper ocean, but there are signals of salinity error at greater depths as well.

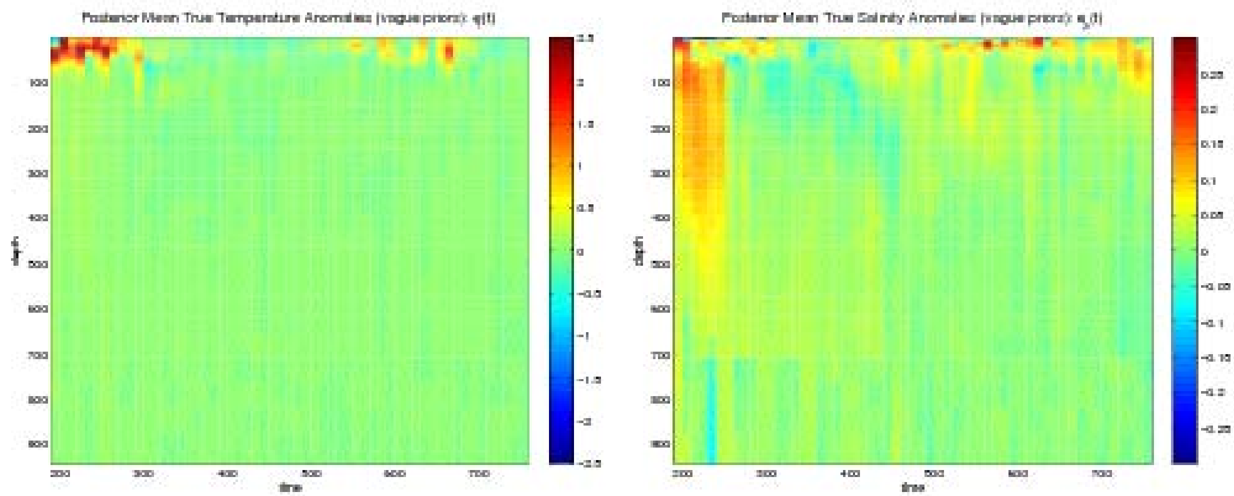

***Figure 4. Depth vs. time profiles of the mean of the posterior distributions for time-dependent error processes, e, from MFS-Error-BHM. Temperature error process profile time series for one of 13 sub-regions is shown the the lefthand panel, and salinity error profiles are shown in the righthand panel. The MFS-Error-BHM output shown here is from a run of the model that does not standardize the misfit and anomaly error in data stage distributions, and prior variances on parameters of the error process model (prior) are vague.***

Background error covariance matrix structures for two snapshots during the MFS forecast period are shown in Fig. 5. Uncertainty in the background error covariance magnitude for each snapshot is quantified by the second moment of the posterior distribution; represented in Fig 5 as the background error covariance standard deviation (righthand panels).

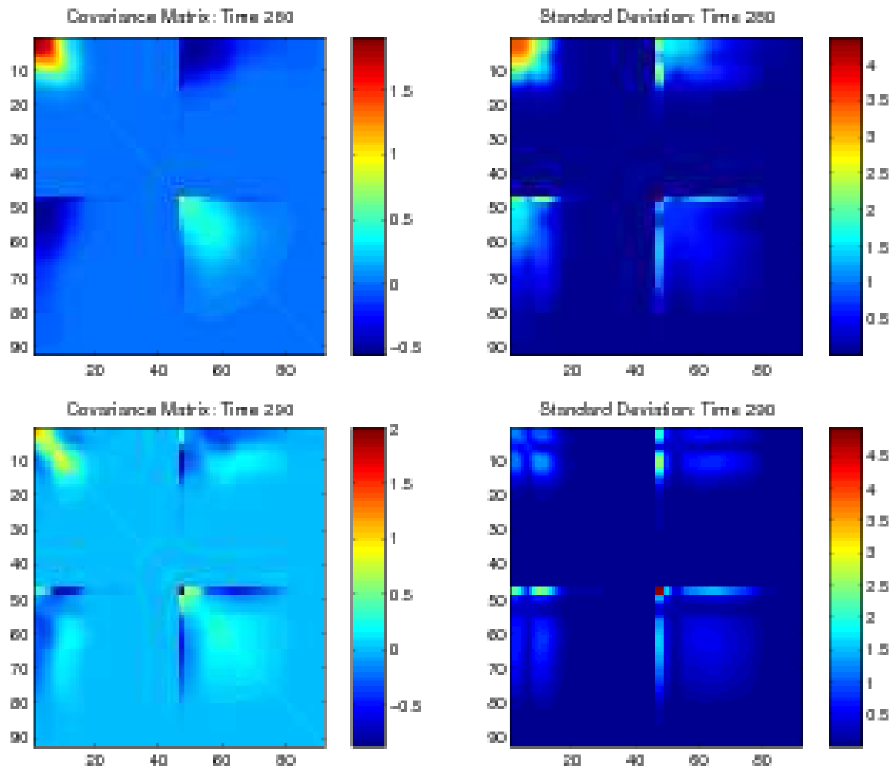


***Figure 5. Contour plots of the MFS background error covariance matrix, B, structures at two forecast times (left panels) demonstrating temporal evolution. Rows and columns of B are comprised of temperature at all levels and salinity at all levels, so that the upper right block is the T,T covariance, the lower left is the S,S covariance, and the off diagonal blocks are T,S covariances. The panels at right depict the spread in the error covariance estimates (i.e. the second mode of the posterior distribution for the background error covariance)***

### *MFS-SuperEnsemble-BHM*

An all-hands meeting was held at NWRA/CoRA in August 2007 to: 1) decide final calculations in the MFS-Wind-BHM and MFS-Error-BHM models as described above; and 2) design experiments for multi-parameter and multi-model applications of MFS-SuperEnsemble-BHM. Professor Di Lorenzo has joined the PI team with responsiblities to direct the development of a second Mediterranean Sea ocean forecast system in MedROMS. While MedROMS development is underway at Georgia Tech, Dr. Paolo Oddo (INGV, Bologna) will begin multi-parameter studies in the MFS research ocean model (NEMO).

The overarching super ensemble methodology, for the multi-parameter and mulit-model experiments, is taken from Berliner and Kim (2007). In this formalism, model integrations for the same target process and forecast period are each used as separate data stages. Observed data can also be combined with the data stage components from the model integrations. A process model is developed from first principles; ideally independent of numerical model results. The identification of appropriate process models forms a critical challenge in this research.

Basin-wide and sub-regional upper ocean salinity has been chosen as the target process for the mulit-parameter experiments in MFS-NEMO. A time-period coinciding with known Levantine Intermediate Water (LIW) formation will be selected for the initial implementations. Issues remaining to be addressed include: dimension reduction (i.e. efficient derivation of target processes and summary fields; identification of specific model bias terms; possible time-dependence in model bias; and the formulation of the process model. Candidate physical processes to be varied in creating the ensemble of MFS-NEMO multi-parameter integrations include: horizontal mixing (momentum and/or tracers); vertical mixing (momentum and/or tracers); etc. Table 2 outlines the number of integrations to be considered in a sample multi-parameter experimental design wherein only the vertical mixing of momentum is varied, and a 10-member ensemble is considered for the MFS-NEMO model only.
A single (perhaps existing) integration in the operational model will be used in this scenario.

***Table 2. Sample experimental design for MFS-SuperEnsemble-BHM multi-parameter experiments, using the research (NEMO) and operational ocean models from MFS. In this sample experiment, the vertical mixing parameterizations for momentum are varied between the K-Profile Parameterization (KPP) and a Pacanowski-Philander parameterization that are both implemented in the MFS models.***

| | | |
|---|---|---|
| *Model and Data Assimilation Systems* | MFS research model (NEMO)<br>MFS operational model | 2 |
| *Physical process* | Vertical mixing | 1 |
| *Parameterizations (i.e. packages)* | KPP<br>Pacanowski-Philander | 2 |
| *Replicates* | 10-member ensemble MFS-NEMO<br>1 existing calculation MFS-Ops | 20<br>1<br>*total of 21 integrations* |

Similar considerations apply in the multi-model case, and Table 3 demonstrates a similar hypothetical experimental design. However, the bias terms for the MedROMS model will be harder to determine since a long forecast record does not exist for that model as it does for MFS. MedROMS forecast model development will depend on cooperation and lessons learned from MFS.

***Table 3. As in Table 2, but for multi-model experiments using MFS-NEMO and MedROMS.***

| | | |
|---|---|---|
| *Model and Data Assimilation Systems* | MFS research model (NEMO)<br>MedROMS | 2 |
| *Physical process* | Vertical mixing | 1 |
| *Parameterizations (i.e. packages)* | KPP<br>Pacanowski-Philander | 2 |
| *Replicates* | 10-member ensembles, both models | 20<br>*total of 40 integrations* |

## RESULTS

- Four versions of MFS-Wind-BHM have been constructed and run through a test forecast period.
- MFS-Error-BHM has been implemented in the Gulf of Lions sub-region of the MFS domain
- logistics for multi-parameter and multi-model MFS-SuperEnsemble-BHM have been scoped.

## RELATED PROJECTS

"Bayesian Hierarchical Models to Augment the Mediterranean Forecast System" (same title), Prof. L. Mark Berliner (Principal Investigator), Department of Statistics, Ohio State University, Grant Number N00014-05-1-0336; Prof. Christopher K. Wikle (Principal Investigator), Department of Statistics, University of Missouri, Grant Number N00014-05-1-0337.

"The Mediterranean Forecast System: Toward Environmental Prediction", Prof. Nadia Pinardi, Scientific Coordinator, Istituto Nazionale di Geofisica e Vulcanologia (National Institute of Geophysics and Volcanology) , EU Contract Number EVK3-CT-2002-00075

"Continued Development of 4D-Variational Data Assimilation and Adjoint-Based Methods of Sensitivity Analysis and Applications Using ROMS", Andrew M. Moore (Principal Investigator), Department of Ocean Sciences, University of California, Santa Cruz, Grant Number N00014-06-1-0406.

## REFERENCES

Berliner, L.M., and Y. Kim, 2007: "Bayesian design and analysis for super ensemble based climate forecasting", J. Climate, in press.

Bonazzi, A., N. Pinardi, R.F. Milliff, C.K. Wikle and L.M. Berliner, 2007: "Ensemble ocean forecast methodology and impacts in the Mediterranean Forecast System", in preparation.

Milliff, R.F., A. Bonazzi, C.K. Wikle, L.M. Berliner and N. Pinardi, 2007: "Ensemble surface winds from Bayesian hierarchical models for the Mediterranean Forecast System", in preparation.

Royle, J.A., L.M. Berliner, C.K. Wikle and R.F. Milliff, 1998: "A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea", *Case Studies in Bayesian Statistics IV,* C Gatsonis, R.E. Kass, B. Carlin, A. Cariquiry, A Gelman, I. Verdinelli and M.West (Eds.), Springer-Verlag, 367-381.

Wikle, C.K., R.F. Milliff, D. Nychka, and L.M. Berliner, 2001: "Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds", *J. Amer. Stat. Assoc.,* **96***,* 382-397.

Wikle, C.K., S. Dobricic, N. Pinardi, R.F. Milliff and L.M. Berliner, "A Bayesian hierarchical model for time-dependent background error covariance in an optimal interpolation data assimilation system for the Mediterranean Forecast System", in preparation.

**PUBLICATIONS** *(see also "in preparation" papers cited above)*

Berliner, L.M., and C.K. Wikle, 2007: "An approximate Importance Sampling Monte Carlo for Data Assimilation", *Physica D*, **230**, 37-49.

Wikle, C.K., and L.M. Berliner, 2007: "A Bayesian Tutorial for Data Assimilation", *Physica D*, **230**, 1-16.